

Title	マルコフモデルを用いたアクセスログ解析
Sub Title	
Author	守谷, 元宏(Moritani, Motohiro) 林, 高樹(Hayashi, Takaki)
Publisher	慶應義塾大学大学院経営管理研究科
Publication year	2016
Jtitle	
JaLC DOI	
Abstract	
Notes	修士学位論文. 2016年度経営学 第3212号
Genre	Thesis or Dissertation
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40003001-00002016-3212

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the Keio Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

慶應義塾大学大学院経営管理研究科修士課程

学位論文（ 2016 年度）

論文題名

マルコフモデルを用いたアクセスログ解析

主 査	林高樹
副 査	大林厚臣
副 査	山本晶
副 査	

氏 名	守谷元宏
-----	------

論文要旨

所属ゼミ	林高樹 研究会	氏名	守谷元宏
(論文題名) マルコフモデルを用いたアクセスログ解析			
(内容の要旨) マーケティングにおいて、一般的にアクセスログは「サイトにどうやって、どのページから入ってきたか」、「目的のページにアクセスされたか」が最も注目されており、それらに対する解析手法は広く認知されている。しかし、それら以外の部分に対しては、十分な活用が行われているとは言い難く、解析も一般的ではない。 そこで本論文では、サイト内を複数のブロックに分け、その加工されたアクセスログに対してマルコフ連鎖モデルを用いて、閲覧者のサイト内の行動について解析を行う。これにより得られたブロック間の推移確率を利用し「コンバージョンページへの推移確率」、「滞留率(閲覧者のページ推移回数とサイト内閲覧継続率)」、「滞留状態における推移回数毎のコンバージョンページへの推移確率」の3つを算出する解析手法を提案する。 これらを休日平日間、月間などの時間で比べることにより、サイト内の回遊状態がどうなっているかが分かり、更には閲覧者の興味の変化や外部環境の変化をいち早く察知する事ができる。これらの情報をもとにサイト内の最適化や改善や、広告戦略などに活用することが可能となる。			

内容

図表目次	5
1. 序論	7
1.1. 本研究の目的	7
1.2. 成果	7
1.3. 用語	8
2. 研究の背景	9
2.1. アクセス解析の現状	9
2.2. 企業におけるアクセスログの扱い	9
3. 関連研究	10
3.1. アクセスログに関する研究	10
3.2. アクセスログに関する研究のマーケティングへの応用	10
4. 解析の手法	11
4.1. アクセスログの加工	11
4.2. アクセスログの整理	11
4.3. アクセスログの解析	11
4.3.1. 初期状態分布	11
4.3.2. 推移確率行列	11
4.3.3. 推移	12
4.3.4. コンバージョンブロックへの推移確率	12
4.3.5. 滞留曲線	13
4.3.6. 滞留中に於けるコンバージョンブロックへの推移確率曲線	13
5. データ解析結果	14
5.1. アクセスログの加工	14
5.2. アクセスログの整理	15
5.3. アクセスログの解析	17
5.3.1. 初期状態分布	17
5.3.2. 推移確率行列	17
5.3.3. コンバージョンブロックへの推移確率	18
5.3.4. 滞留曲線	21
5.3.5. 滞留中に於けるコンバージョンブロックへの推移確率曲線	22
6. 結果及び考察	24
6.1. 初期状態分布と推移確率行列	24
6.2. コンバージョンブロックへの推移確率	27
6.3. 滞留曲線	27
6.4. 滞留中のコンバージョンブロックへの推移確率	28

6.5. コンバージョン改善に向けた提案.....	30
7. 本手法の特徴と限界.....	32
7.1. アクセス解析としての評価.....	32
7.1.1. 他のアクセス解析手法との比較.....	32
7.1.2. 他のアプリケーションとの比較.....	32
7.2. マーケティングに於ける評価.....	33
7.3. 手法の限界と課題.....	33
謝辞.....	34

図表目次

図 1 アクセスログの整理	16
図 2 休日の初期状態	17
図 3 平日の初期状態	17
図 4 休日の推移確率行列	18
図 5 平日の推移確率行列	18
図 6 「AMA」行と END 行	19
図 7 「PRO」と END 行	19
図 8 「Reform_C」と END 行	20
図 9 「SH_C」と END 行	20
図 10 各コンバージョンブロック行と END 行の極限推移確率	21
図 11 休日に於ける滞留曲線	21
図 12 平日に於ける滞留曲線	22
図 13 休日の滞留中に於ける「AMA」への推移確率	22
図 14 休日の滞留中に於ける「PRO」への推移確率	23
図 15 平日の滞留中に於ける「AMA」への推移確率	23
図 16 平日の滞留中に於ける「PRO」への推移確率	23
図 17 初期状態分布の差分	24
図 18 推移確率行列の差分	25
図 19 休日に強く見られる推移	25
図 20 平日に強く見られる推移	25
図 21 各商品の自ブロックへの推移確率	26
図 22 「Fire」ページと推移確率の差分	26
図 23 「Fabric」ページと推移確率の差分	27
図 24 各コンバージョンの差分	27
図 25 滞留曲線の比較	28
図 26 滞留中に於ける「AMA」への推移確率曲線の比較	29
図 27 滞留中に於ける「PRO」への推移確率曲線の比較	30
図 28 ゴールフロー [17]の動画の一部	33
表 1 X 社に於けるブロック分け	14
表 2 離脱率による比較	28
表 3 「AMA」への推移確率曲線の推移回数による比較	29
表 4 「PRO」への推移確率曲線の推移回数による比較	30
数式 1 初期状態分布 $S(0)$	11
数式 2 推移行列 a	12

数式 3 推移確率行列 P	12
数式 4 推移計算	12
数式 5 吸収構造を持つ推移確率行列 P^*	13
数式 6 滞留中に於けるコンバージョンブロックへの推移確率 P^{**}	13

1. 序論

1.1. 本研究の目的

近年、インターネットの普及により、どの企業も当たり前自社のホームページを持つようになった。それに伴い、アクセスログの解析(以下、アクセス解析)が発展、普及してきているが、現在のアクセス解析において重視されているのは、閲覧者向けに社外のサイトに存在している広告の種類(リファラ解析)や、それらの広告群がどの様に閲覧者のアクセスに貢献しているか(アトリビューション分析)、または閲覧者が探索エンジンから自社のサイトをどのような言葉で探索してアクセスしているか(SEO)などである。これらのアクセス解析では、閲覧者がサイトに推移してきた最初のページ(以下、ランディングページ)、どこを経由したか(以下、リファラ)、コンバージョンページに至るまでの経路に注力しており、サイト内での閲覧者の全体の動きを重要視する解析手法について近年のマーケティングに於いては触れられていない。閲覧者全体の動きについての解析は学術的には 1996 年の Bestavors [1] を始めとして様々なもの [2] [3]があるが、それらがマーケティングに応用されているとは言い難い。

そこで本研究ではサイト内に於ける閲覧者の動き全体に対して解析を行い、近年のマーケティングでは使われていない視点により、アクセスログの新たな価値を再確認する。具体的には従来のアクセスログ研究でも使われているマルコフモデルを用いて、サイト内での各ユーザーのアクセスログを解析する。その際、閲覧者のアクセスログを個別のブロックに分けてから、マルコフ連鎖を作り、計算することで、閲覧者のサイト内の動きを可視化する。

結果、アクセスログ解析から閲覧者の動向や興味、意図しているページ間の推移の有無が推定され、それらからサイト内の最適化のヒントや消費者ニーズを知る事が出来る事が期待される。

1.2. 成果

上記の手法にて、株式会社 X(以下、X 社)から提供された 2016 年 5 月 1 日、3 日から 31 日の合計 30 日間の X 社ホームページへのアクセス解析を統計ソフト R [4]で行う。まず、ブロックからブロックの推移を確率的に算出し、それらを休日平日間で比べることで閲覧者の推移の違い、初期状態分布の違いを確認する。次に得られたブロック間の推移確率を利用し「コンバージョンページへの推移確率」、「滞留曲線(閲覧者のページ推移回数とサイト内閲覧継続率)」、「滞留状態における推移回数毎のコンバージョンページへの推移確率」の 3 つを算出する事。さらにサイト構造についての改善点や閲覧者の興味を観察する。

1.3. 用語

本論文内において重要な用語の解説をする。

アクセスログ

アクセスログはサーバーの種類により、種類が別れる。今回使用したのは Apache の結合ログ形式なのである。

アクセス解析

一般的にはメンテナンスを目的としたアクセス解析とマーケティングを目的としたアクセス解析があるが、本論文においては後者を指す。

コンバージョン

商品購入や資料請求など、閲覧者がサイト運営の意図するページに推移、または意図する動作を行い、閲覧者から顧客に「転換」すること。

ランディングページ

閲覧者がそのサイト内に入ってきた最初のページ。トップページであるとは限らない。

リファラ

閲覧者がそのサイト内に入ってきた時に経由した媒体。広告であったり、検索エンジンであったりする。

2. 研究の背景

2.1. アクセス解析の現状

アクセス解析ソフトは様々なものがある。有名な例としては無料で使える Google Analytics [5] がある。これはサーバーに関する詳しい知識がなくとも使え、高度な解析を行えるが、アクセスログ自体は Google 社の所有物になってしまうため、コンプライアンス上導入出来ない企業が存在する。

一方で、所有するアクセスログに対する解析ソフトにも様々なもの [6] [7] があるが、高度な解析を行うためにはサーバーやプログラミングに対する知識が必要であり、ある程度の IT スキルが必要である。

アクセス解析の手法は目的に合わせて様々なものがある。単純にサイト内の訪問者数、Cookie を使うことで新規ユーザーやリピーターの割合、ページビュー数、滞在時間、直帰率(アクセス後すぐにサイトから出てしまう確率)なども期間を決めて調べればアクセス解析といえる。より複雑なものでは閲覧者が検索エンジンでどんな言葉を使ってアクセスしてきたかを調べる検索エンジン最適化(SEO, Search Engine Optimization)、どの広告枠からアクセスしてきたかを調べるリファラ解析、広告群がどの様に閲覧者のアクセスに貢献したかを調べるアトリビューション分析などがある。各分析は広告費の宣伝効果を算出するためによく使われているため、基本的にはランディングページ付近を解析している。

サイト内の解析に関しては、サイト内検索で使われた言葉を分析するサイトサーチアナリティクスがある。これは直接的なマーケティングというよりも、サイト内の改善を行い、顧客満足度を上げてコンバージョンを目指すアクセス解析であると言われている [8]。

2.2. 企業におけるアクセスログの扱い

調査 [9] では、2015 年 12 月時点で上場企業 3604 社中 2806 社(77.9%)が Google Analytics を導入済みであるが、全てのページに導入しているわけではなかった。また、最新版の Google Analytics タグを導入している企業は 1364 社(37.8%)であった。別の調査 [10] ではデータログを取得・管理している企業は 332 社中 148 社(44.6%)であり、その中でもアクセスログを取得・管理しているのは 79 社(全体の 23.8%)のみであった。

これらの調査では傾向として従業員規模が大きくなるにつれてデータログを取得・管理している傾向が強くなり、規模が小さくなるにつれて取得・管理していない傾向が強くなることから分かる。また、ログを管理しない理由としては「ログを収集しても分析・活用方法が分からない」、「規模が小さいため不要」、「ログ管理に関して知識があまりない」が上位であった。

3. 関連研究

3.1. アクセスログに関する研究

アクセス解析に関する研究は数多くある。例えば、Bestavors [1]や Zukerman [2]はマルコフモデルを用いて閲覧者の推移するページの予測を行った。他にも Montgomery [3]は解析の際にサイト内のページの分類を行うことで、解析結果を理解しやすくしている。また、ページ間の推移関係性の強さに関するもの [11] [12]も存在している。近年では、これらの手法を分類法、モデルごとに適応し、どのように予測精度が変わるかについて研究したものもある [13]。

3.2. アクセスログに関する研究のマーケティングへの応用

マーケティングに於いてアクセスログは宣伝費とその効果の測定のために、閲覧者のランディングページ付近、閲覧者の流入元であるリファラ、閲覧者が顧客に変わるコンバージョンの 3 つが重視されている。マーケティングに於けるアクセス解析の歴史では、当初は閲覧者が流入してくるに経由した広告のみを重視するラストクリック型であったが、2010 年頃から、それ以外の目に映ったが流入に直接的に関係していない間接的な広告への効果測定を重視し始め「アトリビューション分析」という用語が使われ始めているが、その手法や範囲などは分析者ごとに違い、漠然とした用法になっている [14] [15] [16]。

4. 解析の手法

本手法ではアクセスログデータの加工、整理を行ってから、解析を行う。解析により「コンバージョンページへの推移確率」、「滞留曲線」、「滞留中のコンバージョンページへの推移確率」を知ることが出来る。これらの結果を時間ごと、休日平日間や月間で比べることで閲覧者の行動を知ることが出来る。

4.1. アクセスログの加工

後述するように推移確率行列を使ってアクセスログを解析するため、サイト内のページ全てを個別に解析してしまうと膨大な要素を持つ行列になる。そこで Montgomery [3]らが行った様に、サイトのブロック化を行い、その要素数を抑えたとともに、注目したいサイト内のブロック(コンバージョンブロック)を決める。

4.2. アクセスログの整理

アクセスログに残されたサイトの閲覧者の情報には人間以外の情報収集用ボットや1ページだけ見てすぐに帰ってしまう閲覧者(情報収集用ボットなど)が含まれている。このままの状態では、解析結果にノイズが含まれてしまうため、アクセスログを整理する必要がある。予想されるそのサイトの閲覧者の行動を考え、それに含まれない閲覧者の情報を削除する。

4.3. アクセスログの解析

4.3.1. 初期状態分布

下記のように各閲覧者のランディングブロックのみを累計した初期状態行列 $S_{(0)}$ を計算する。初期状態行列 $S_{(0)}$ は $1 \times n$ の行列を持つ。

$$S_{(0)} = \left[\frac{V_1}{\sum_{k=1}^n V_k}, \frac{V_2}{\sum_{k=1}^n V_k}, \dots, \frac{V_n}{\sum_{k=1}^n V_k} \right], \quad \sum_{h=1}^n \frac{V_h}{\sum_{k=1}^n V_k} = 1$$

数式 1 初期状態分布 $S_{(0)}$

4.3.2. 推移確率行列

アクセスログを各閲覧者に分け、ブロック間の移動をマルコフ連鎖モデルで作られた推移行列 α で表す。推移行列 α は、 $n \times n$ 行列を持つ。その後、ブロック間の推移に対応した行(始点)と列(終点)が交わる場所をカウントしていく。

例えば、閲覧者が「Toppage」→「Bath」→「Catalog」→「END」と推移した場合は、「Toppage」→「Bath」、「Bath」→「Catalog」、「Catalog」→「END」の 3 回の推移があると考え、Toppage 行 Bath 列、Bath 行 Catalog 列、Catalog 行 END 列をカウントする。END 行 (サイトからの離脱行)は性質上、累計が必ず 0 になる。

$$a = \begin{bmatrix} V_{1,1} & \cdots & V_{1,n} \\ \vdots & \ddots & \vdots \\ V_{n,1} & \cdots & V_{n,n} \end{bmatrix}$$

数式 2 推移行列 a

この推移行列の各行のベクトル総数 ($V_1 = V_{11} + V_{12} + \cdots + V_{1n}$) で各行のベクトル ($V_{11}, V_{12}, \cdots, V_{1n}$) を割ることで推移確率を算出し、各ブロック間の推移確率行列 P を作る。

$$P = \begin{bmatrix} \frac{V_{1,1}}{V_1} & \cdots & \frac{V_{1,n}}{V_1} \\ \vdots & \ddots & \vdots \\ \frac{V_{n,1}}{V_n} & \cdots & \frac{V_{n,n}}{V_n} \end{bmatrix}, \quad \sum_{k=1}^n V_{n,k} = V_n$$

数式 3 推移確率行列 P

4.3.3. 推移

上記の初期状態行列 $S_{(0)}$ に対して、推移確率行列 P を掛けることで、推移後の状態 $S_{(1)}$ を得ることが出来る。これを $k+1$ 回繰り返した状態は以下の式により得られる。

$$S_{(K+1)} = S_{(K)} P = \cdots S_{(0)} P$$

数式 4 推移計算

4.3.4. コンバージョンブロックへの推移確率

数式 4 では単純に閲覧者の推移を確率的に追っていただけであり、本研究の目的には不十分である。そこで推移確率行列 P のコンバージョンブロックの行と離脱の END 行を吸収構造にすることで

推移を固定する推移確率行列 P^* を作る。推移確率行列 P^* を十分な回数 $S_{(0)}$ に掛けることで、閲覧者がコンバージョンブロックに推移する確率(極限推移確率状態)が算出することが出来る。(コンバージョンブロックを複数設定した場合、実際には閲覧者がコンバージョンブロック A からコンバージョンブロック B に推移していても、コンバージョンブロック A で吸収してしまうため、コンバージョンブロック B はカウントされず、コンバージョンブロック A しかカウントされない)。

$$P^* = \begin{bmatrix} \frac{V_{1,1}}{V_1} & \dots & \frac{V_{1,n}}{V_1} \\ \vdots & \ddots & \vdots \\ 0 & 1 & 0 \\ \frac{V_{n,1}}{V_n} & \dots & \frac{V_{n,n}}{V_n} \end{bmatrix}$$

数式 5 吸収構造を持つ推移確率行列 P^*

4.3.5. 滞留曲線

閲覧者はブロック推移を重ねることで、徐々に離脱していくが、その推移回数と離脱割合の関係をアクセスログから求める。

END 行のみを吸収構造にした推移確率行列 P^* を利用し、初期状態分布 $S_{(0)}$ と掛け合わせる。全体の 99.99%が END 行に吸収される状態になるまで繰り返す。また、END 行に吸収されていない状態を滞留状態とする。

各試行中における $1 - \frac{V_{END}}{V}$ を滞留率とし、Y 軸に滞留率、X 軸に試行回数を取る折れ線グラフを滞留曲線とする。

4.3.6. 滞留中に於けるコンバージョンブロックへの推移確率曲線

初期状態分布 $S_{(0)}$ と推移確率行列のコンバージョン列 P^{**} から滞留中に於けるコンバージョンブロックへの推移確率 P^{**} を算出し、Y 軸に滞留中の推移確率、X 軸に試行回数を取る折れ線グラフを滞留中における推移確率曲線を作成する。

$$P^{**}(k) = \frac{S_{(k+1)conv} - S_{(k)conv}}{1 - S_{(k)END} - S_{(k)conv}}$$

数式 6 滞留中に於けるコンバージョンブロックへの推移確率 P^{**}

5. データ解析結果

実際に本手法を用いて、X社の5月分のアクセスログを休日(1日,3日,4日,5日,7日,8日,14日,15日,21日,22日,28日,29日)と平日(6日,9日,10日,11日,12日,13日,16日,17日,18日,19日,20日,23日,24日,25日,26日,27日,30日,31日)に分けて解析を行った。なお、5月2日分についてはデータが破損していたため、解析には含んでいない。今回使用したデータではユニークユーザー数は3,188,007人、推移回数が71,763,736回であった。

5.1. アクセスログの加工

Montgomery [3]らが行った様に、サイトのブロック化を行った。X社サイト内を30ブロックに分け、各セッションの終わりを示すために「END」を追加し、計31ブロックに分けた。

サイトブロックの内訳は以下の表1のとおりであり、これはX社が社内で使用しているブロック分けを基準にして作られている。

表 1 X社に於けるブロック分け

ブロック名	内容
Apaman	アパート・マンション
Bath	バス
Biz	ビジネス関連
Catalog	カタログ一般
AMA	施工主向けカタログ請求画面
PRO	工務店向けカタログ請求画面
Constr	建築関連
Constructor	新築関連
Door	門・扉
Exter	エクステリア
Fabric	カーテンなどの布製品
Faq	FAQ
Fire	防火戸
Inter	インテリア
Kitchen	キッチン
Lavatory	化粧台
Lineup_TOP	ラインナップのトップページ
Info	企業情報
Public	公共施設

Reform_C	リフォームコンバージョン
Reform_N	リフォーム一般
SH_C	ショールームコンバージョン
SH_N	ショールーム一般
Solar	ソーラーパネル
Support	サポート
Tile	タイル
Toilet	トイレ
Toppage	トップページ
Water	水回り
Window	窓・サッシ

X社の考えるコンバージョンブロックは「AMA」、「PRO」、「Reform_C」、「SH_C」である。

5.2. アクセスログの整理

本研究ではエージェント情報からボットのアクセスを推定し、排除した。また、正常な閲覧者(滞在時間が5秒以上10800秒(3時間)以下)を想定することで直帰した閲覧者や滞在時間が極端に長い、もしくは短い閲覧者を削除した。そのため、アクセスログ内のユニークユーザーは3,188,007人(休日1,079,524人、平日2,108,483人)から1,394,597人(休日464,699人、平日929,898人)まで減少したが、休日平日の構成比は殆ど変わっていない(図1)。

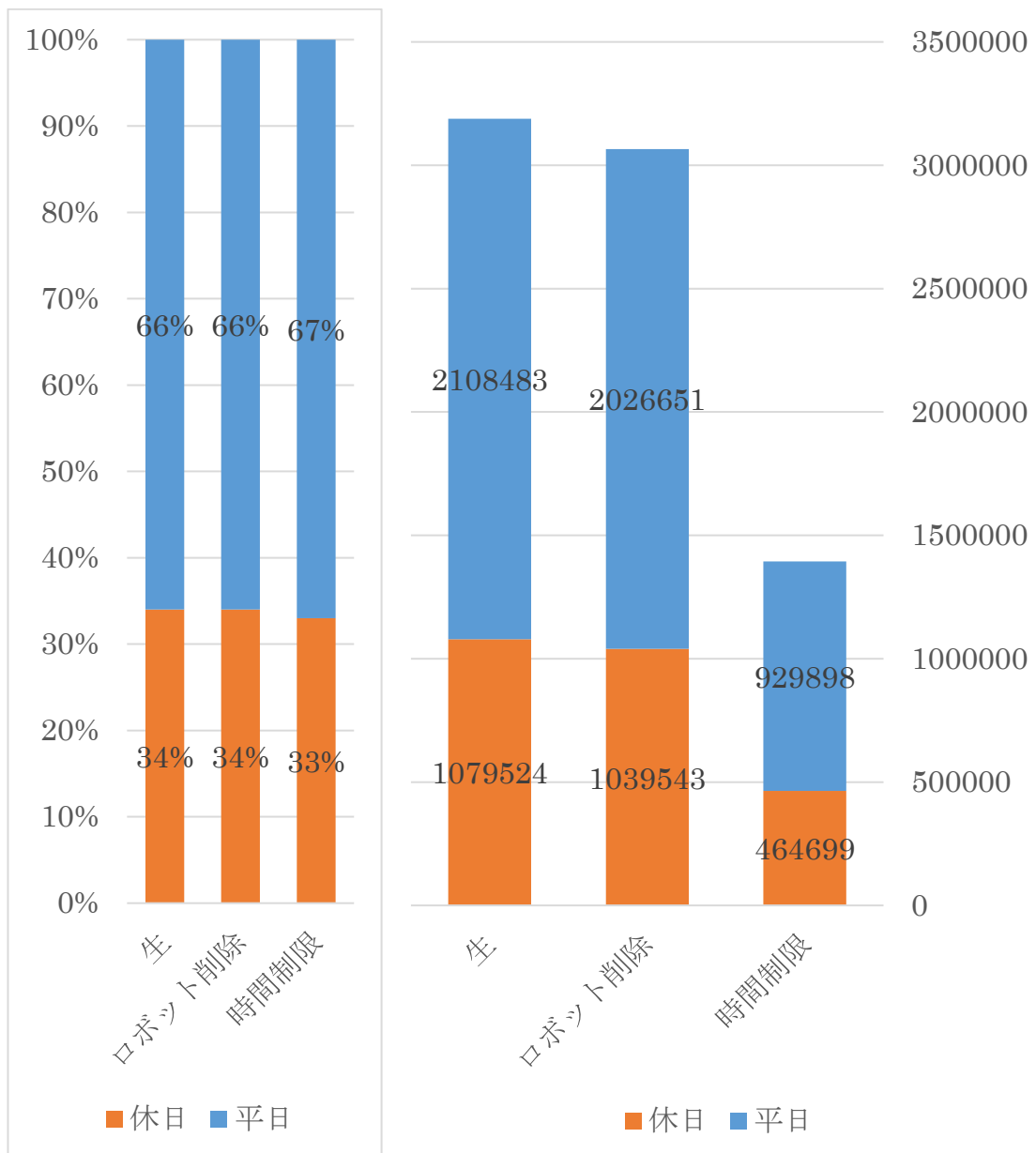


図 1 アクセスログの整理

5.3. アクセスログの解析

5.3.1. 初期状態分布

X社に於いて、初期状態分布 $S_{(0)}$ はサイト内のブロック分けにより、 1×31 の行列になった(図2, 図3)。

ブロック名	Apaman	Bath	Biz	Catalog	AMA	PRO	Constr
割合	0%	6%	1%	1%	0%	1%	0%
Constructor	Door	END	Exter	Fabric	Faq	Fire	Inter
0%	5%	0%	15%	0%	1%	0%	2%
Kitchen	Lavatory	Lineup_TOP	Info	Public	Reform_C	Reform_N	SH_C
8%	2%	9%	2%	0%	0%	4%	0%
SH_N	Solar	Support	Tile	Toilet	Toppage	Water	Window
7%	0%	5%	2%	4%	20%	1%	3%

図2 休日の初期状態

ブロック名	Apaman	Bath	Biz	Catalog	AMA	PRO	Constr
割合	0%	5%	1%	2%	0%	1%	0%
Constructor	Door	END	Exter	Fabric	Faq	Fire	Inter
0%	5%	0%	14%	0%	1%	0%	3%
Kitchen	Lavatory	Lineup_TOP	Info	Public	Reform_C	Reform_N	SH_C
7%	2%	10%	3%	0%	0%	4%	0%
SH_N	Solar	Support	Tile	Toilet	Toppage	Water	Window
4%	0%	4%	2%	4%	24%	1%	3%

図3 平日の初期状態

5.3.2. 推移確率行列

X社に於いて、推移行列はサイト内のブロック分けにより、 31×31 の行列になり、休日の推移確率行列Pは図4、平日の推移確率行列Pは図5になった。

行列P*を4つ作成した。なお、紙面の都合により平日の図は省略する。

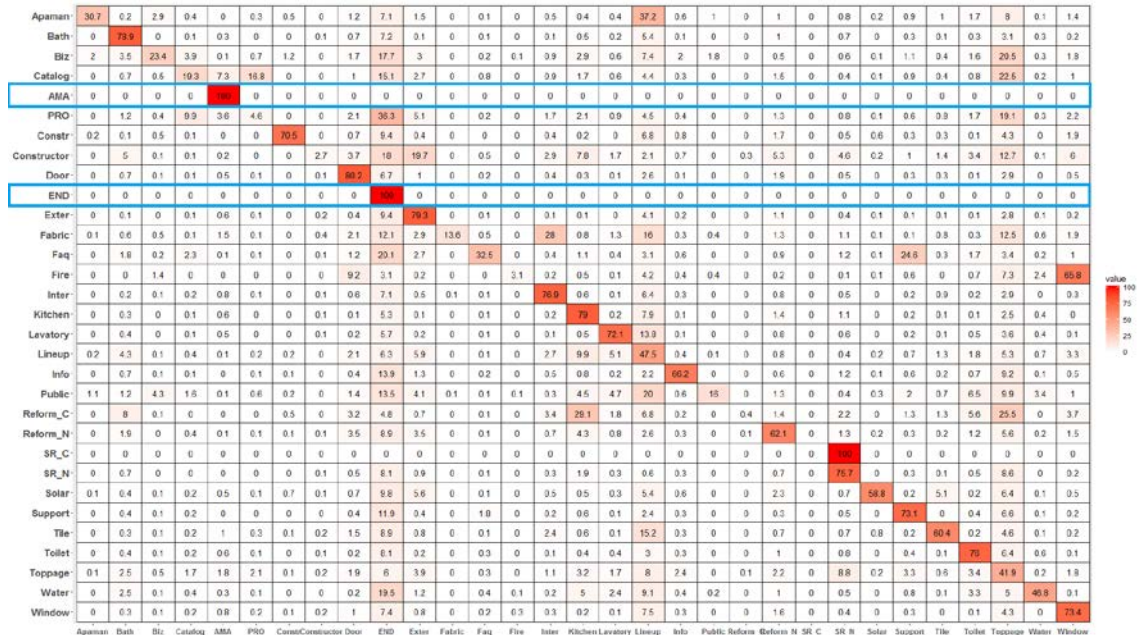


図 6 「AMA」行と END 行

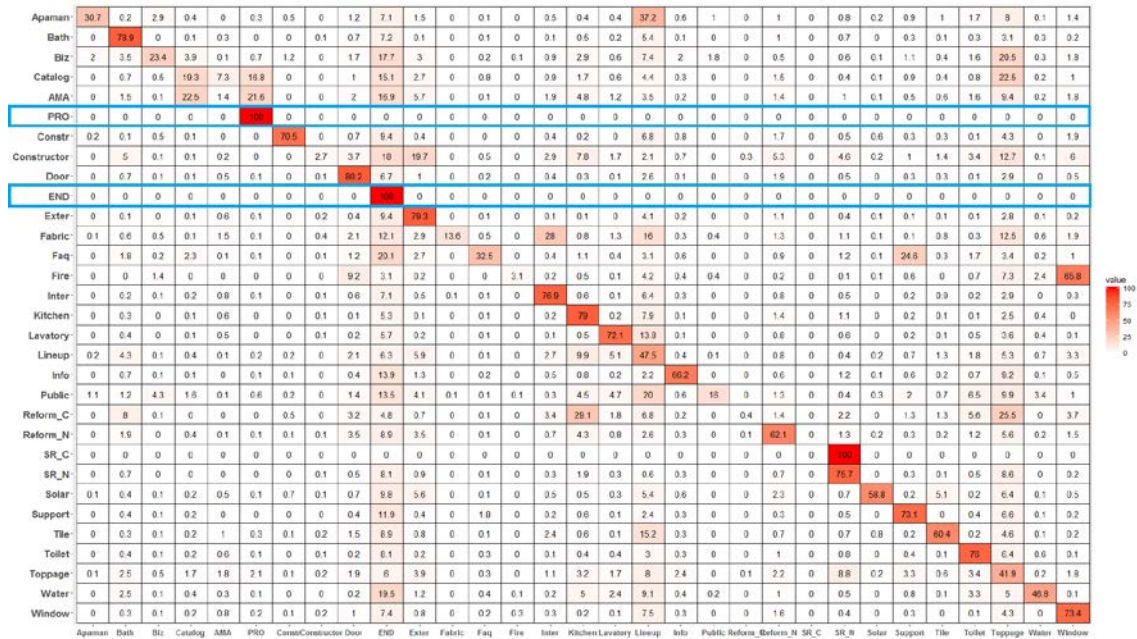


図 7 「PRO」と END 行

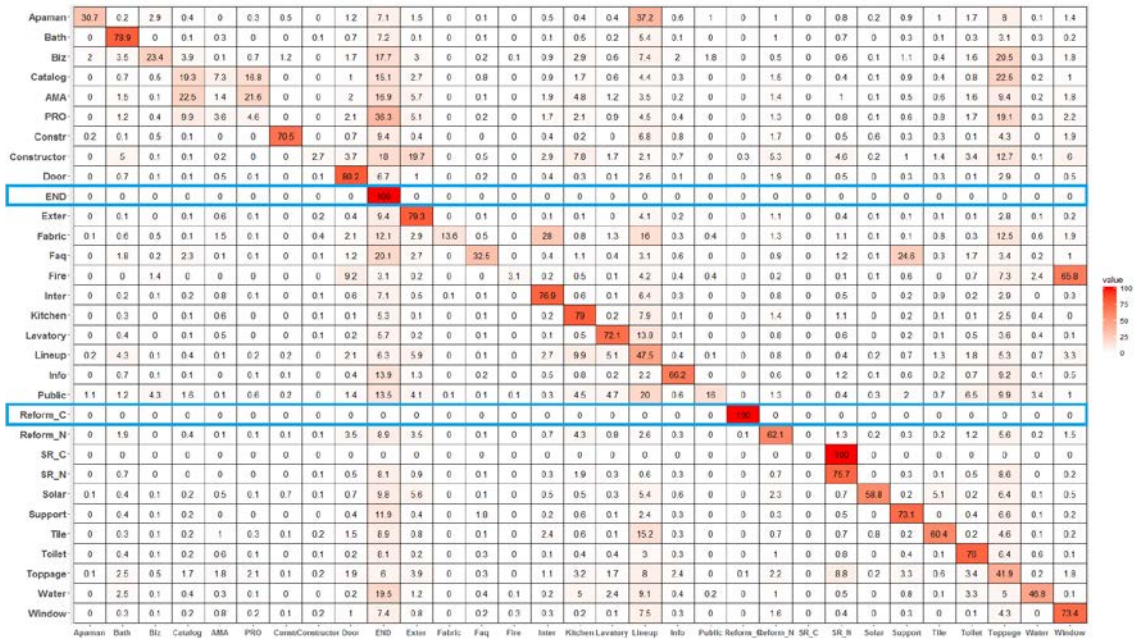


図 8 「Reform_C」と END 行

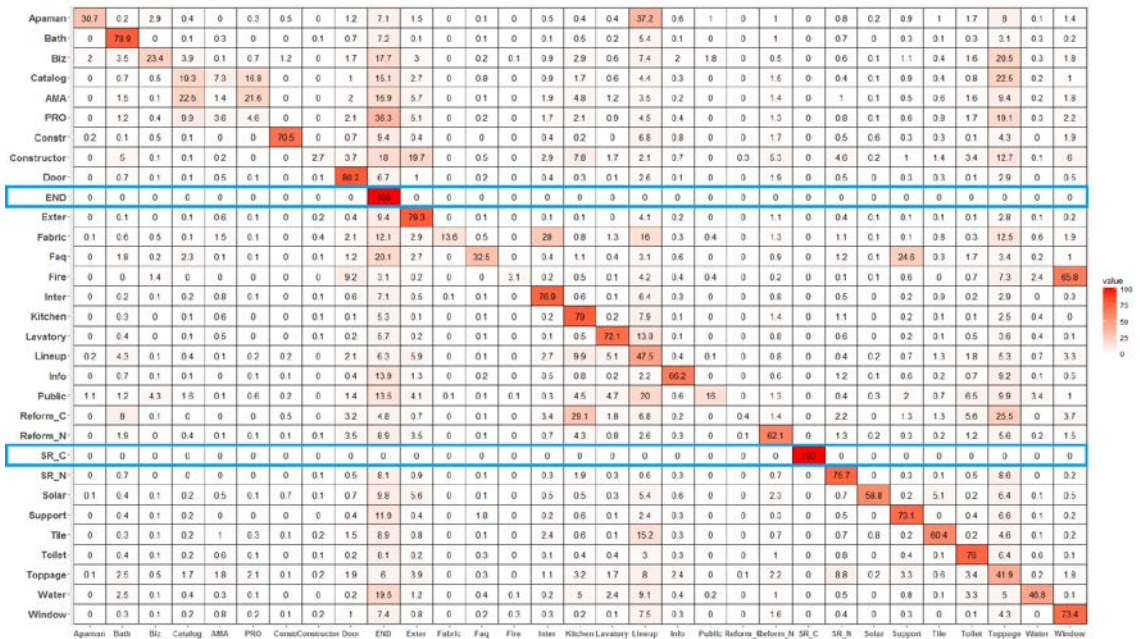


図 9 「SH_C」と END 行

これらと初期状態分布 $s_{(0)}$ を掛け合わせて、各コンバージョンブロック行と「END」行に 99.99%が吸収されるまで繰り返し、極限推移確率を求めた(図 10)。また、見やすくするため、図 10 では各コンバージョンブロック行と「END」行のみを表示した。

AMA	END
7.4%	92.6%
PRO	END
7.2%	92.8%
Reform_C	END
0.4%	99.6%
SH_C	END
0%	100.0%

図 10 各コンバージョンブロック行と END 行の極限推移確率

「Reform_C」, 「SH_C」に対する推移確率は低く、以降の解析には適さないと判断し、以下の解析は「AMA」, 「PRO」の2つのみを解析対象とした。

5.3.4. 滞留曲線

「END」行を吸収構造にした推移確率行列 P^* と初期状態分布 $S_{(0)}$ と掛け合わせ、全体の99.99%が「END」行に吸収されるまで繰り返す、その各試行中の $1 - \frac{V_{END}}{V}$ を滞留率とした。図 11、図 12 の Y 軸は滞留率、X 軸は推移回数である。

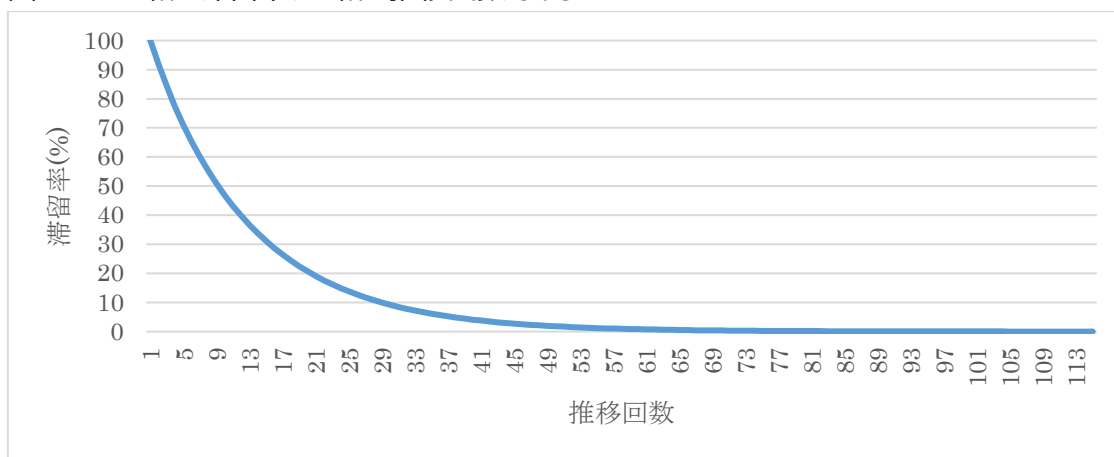


図 11 休日に於ける滞留曲線

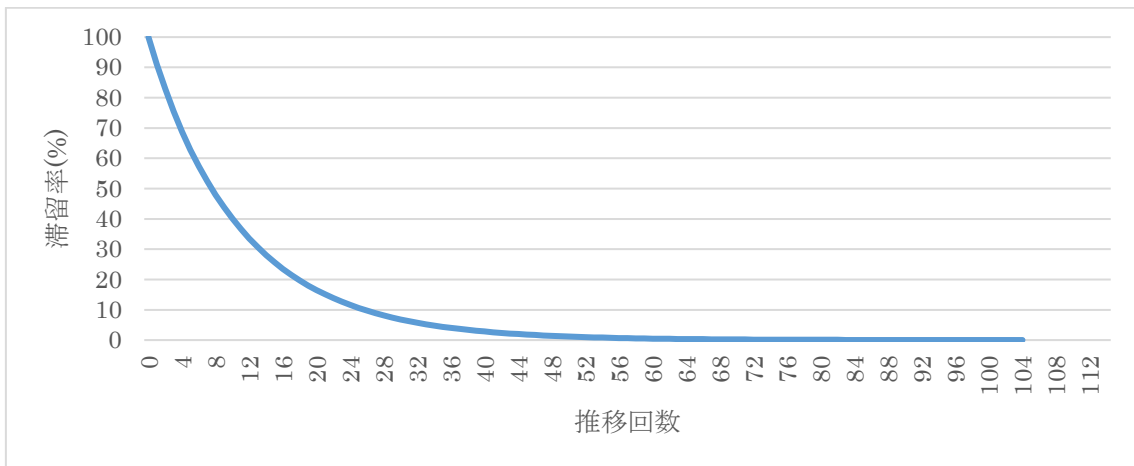


図 12 平日に於ける滞留曲線

5.3.5. 滞留中に於けるコンバージョンブロックへの推移確率曲線

図 5, 図 6 を利用して、滞留中に於ける「AMA」, 「PRO」への推移確率曲線を算出した(図 13, 図 14, 図 15, 図 16)。Y 軸は滞留中の推移確率、X 軸は試行回数であり、この Y 軸の推移確率は全体のアクセス数に対する割合である。

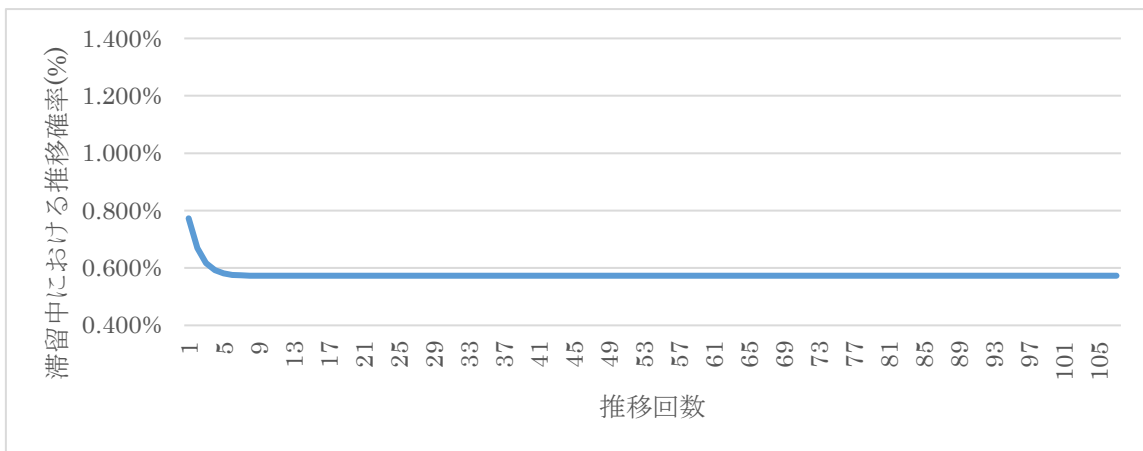


図 13 休日の滞留中に於ける「AMA」への推移確率

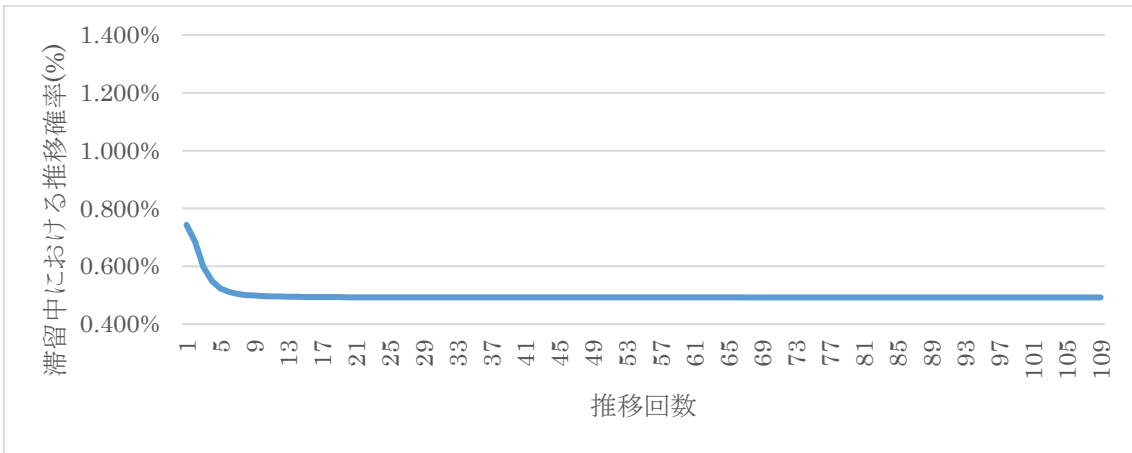


図 14 休日の滞留中に於ける「PRO」への推移確率

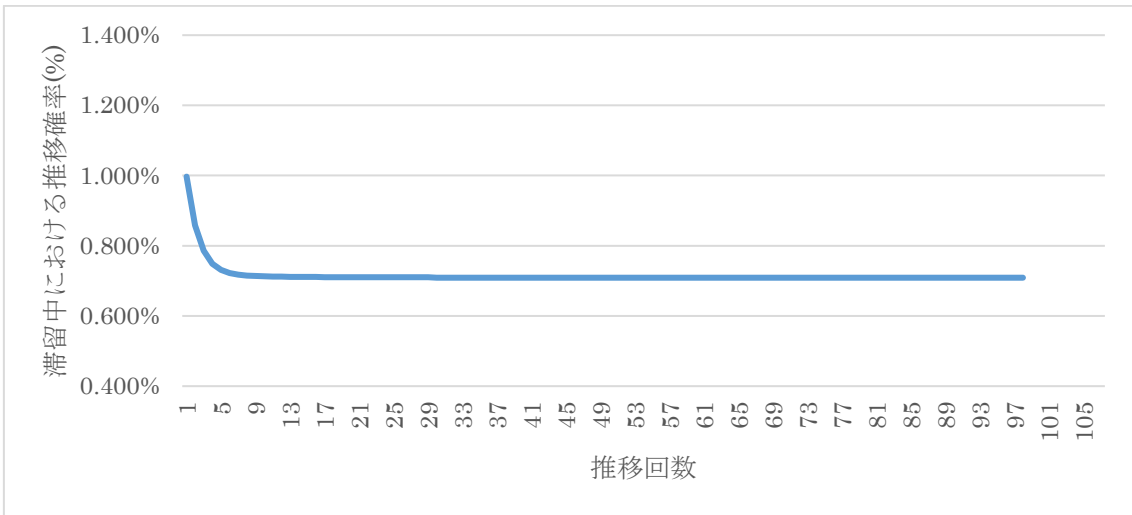


図 15 平日の滞留中に於ける「AMA」への推移確率

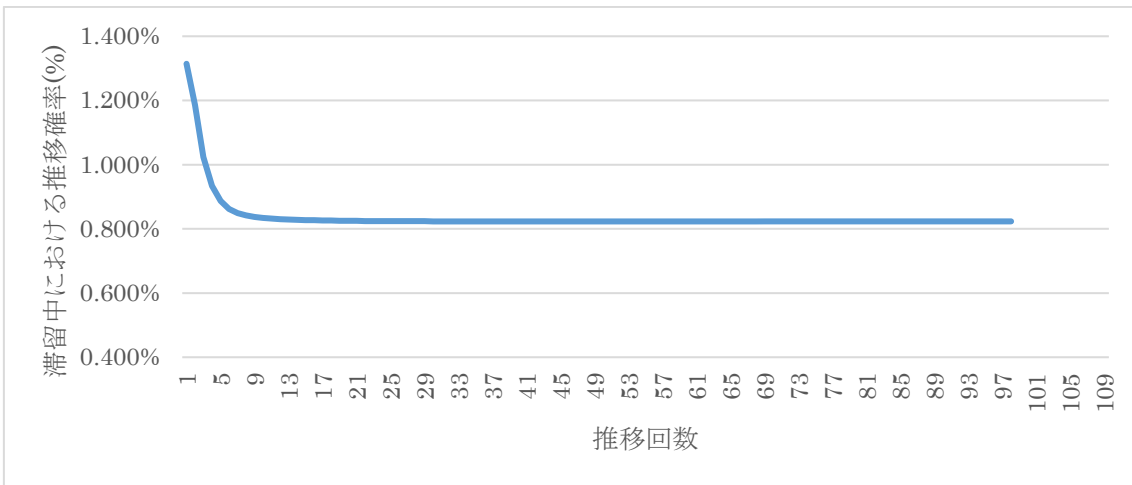


図 16 平日の滞留中に於ける「PRO」への推移確率

6. 結果及び考察

休日平日別に解析したアクセスログの結果をマーケティングやサイト内最適化の視点で比較してみる。今回 X 社より得られたアクセスログは 5 月分だけであるため、この差異が恒常的な差異なのか、それとも季節的な差異なのか、外部環境の変化による差異なのかは不明である。より正しい考察を得るためには、より長期間のアクセスログが必要になる。

また、ユニークユーザー数から、休日に於ける 1%は約 4647 人であり、平日に於ける 1%は約 9299 人であるため、1%の差異だとしても大きい。

6.1. 初期状態分布と推移確率行列

図 2,図 3 を比較してみたのが図 17 である。色が赤いほど休日に推移が多く、青いほど平日に推移が多い。その差は休日では最大 2.7%であり、平日は 4.1%であった。休日は「SH_N」、「Exter」から入ってくる閲覧者が多く、平日は「Toppage」からのサイト内に入ってくる閲覧者が多い事が分かる。

ブロック名	Apaman	Bath	Biz	Catalog	AMA	PRO	Constr
平日ー休日	0.0%	-0.9%	0.2%	0.6%	0.0%	0.5%	0.0%
Constructor	Door	END	Exter	Fabric	Faq	Fire	Inter
0.0%	-0.1%	0.0%	-1.5%	0.0%	-0.1%	0.0%	0.1%
Kitchen	Lavatory	Lineup_TOP	Info	Public	Reform_C	Reform_N	SH_C
-1.0%	-0.3%	1.1%	0.7%	0.0%	0.0%	-0.4%	0.0%
SH_N	Solar	Support	Tile	Toilet	Toppage	Water	Window
-2.7%	0.0%	-1.0%	0.3%	0.0%	4.1%	0.0%	0.1%

図 17 初期状態分布の差分

図 4,図 5 を比較してみたのが図 18 である。「SR_C」は月間数回しかないため、強く影響してしまうので、この図では除いている。色が赤いほど休日に推移が多く、青いほど平日に推移が多い。その差は休日では最大 5.1%であり、平日は 4.3%であった。

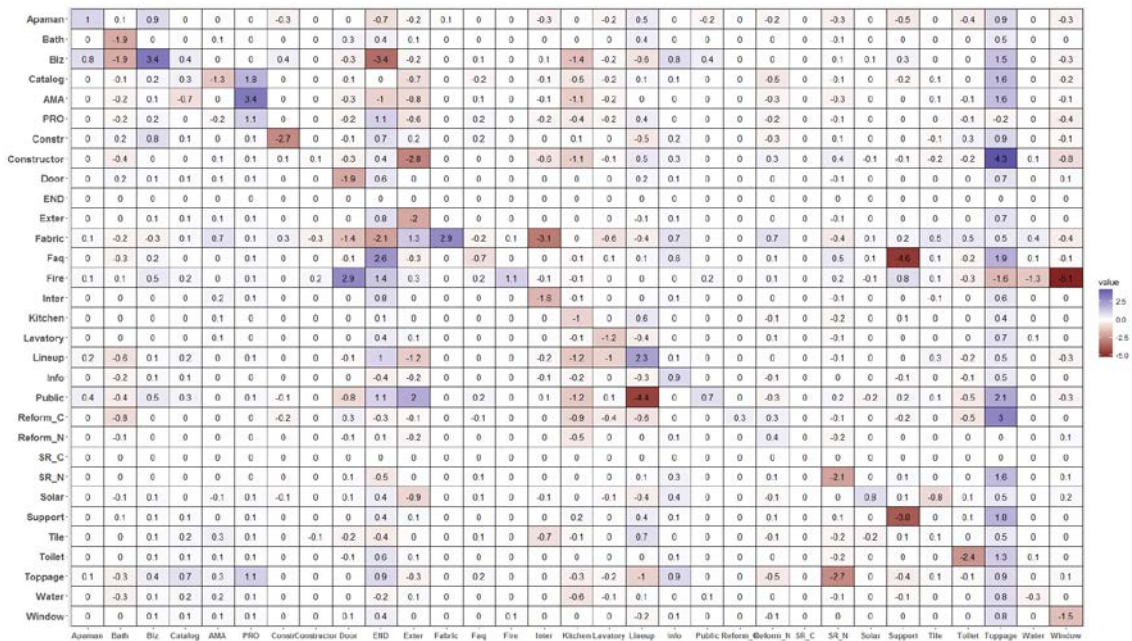


図 18 推移確率行列の差分

休日では「FAQ」から「Support」への推移、「Support」間の推移、「Public」から「Lineup_Top」の推移、「Biz」からの「END」への離脱が比較強く見られた(図 19)。休日のアクセスということで、施工主の推移が多いのではないかと考えられる。

	Support		Support		Lineup_Top		END
FAQ	-4.6%	Support	-3.8%	Public	-4.4%	Biz	-3.4%

図 19 休日に強く見られる推移

平日ではどのコンバージョンブロックからも「Toppage」、「Biz」、「info」、「PRO」に推移する閲覧が多い。また、「FAQ」からの「END」への離脱、「AMA」から「PRO」の推移、「Biz」間の推移が多い(図 20)。平日、アクセスされているブロックから、工務店からの推移が多いのではないかと考えられる。

	END		PRO		Biz
FAQ	2.6%	AMA	3.4%	Biz	3.4%

図 20 平日に強く見られる推移

これらの結果を踏まえて、休日を施工主からのアクセス、平日は工務店からのアクセスと仮定して、各商品ブロックのループ確率(図 21)を見てみると、休日(施工主)は風呂、建築関連、門扉、エクステリア、インテリア、キッチン、化粧台、トイレ、水回り、窓・サッシが多く、平日(工務店)は布製品、防火戸、公共施設、ソーラーパネルが多い事がわかる。これは X 社が施工主に対して「水回り品くらい

しか自身で選択しない、興味を持っていない」という認識と異なる結果になったため、新しいマーケティングの視点になると考えられる。

ブロック名	Bath	Constr	Door	Exter	Fabric	Fire	Inter
ループ確率	-1.9%	-2.7%	-1.9%	-2.0%	2.9%	1.1%	-1.6%
Kitchen	Lavatory	Public	Solar	Tile	Toilet	Water	Window
-1.0%	-1.2%	0.7%	0.8%	0.1%	-2.4%	-0.3%	-1.5%

図 21 各商品の自ブロックへの推移確率

また、休日平日間で推移先の確率に大きな違いが出た「Fire」と「Fabric」に注目してみる。図 22 から分かるように「Fire」トップページ内での推移先は基本的には「Window」か「Door」であり、その推移確率が休日と平日で大きく違う。これを休日 = 施工主、平日 = 工務店と考えると、施工主は防火窓に興味があり、工務店は防火ドアに興味があると考えられる。



図 22 「Fire」ページと推移確率の差分

同様に図 23 の「Fabric」トップページでは「AMA」、「Inter」、「Fabric」であり、推移確率が変わってきている。これも施工主はハニカムスクリーンに興味があり、工務店はオーダーカーテン、ブラインド、ロールスクリーン、ウッドブラインドに興味があると考えられる。

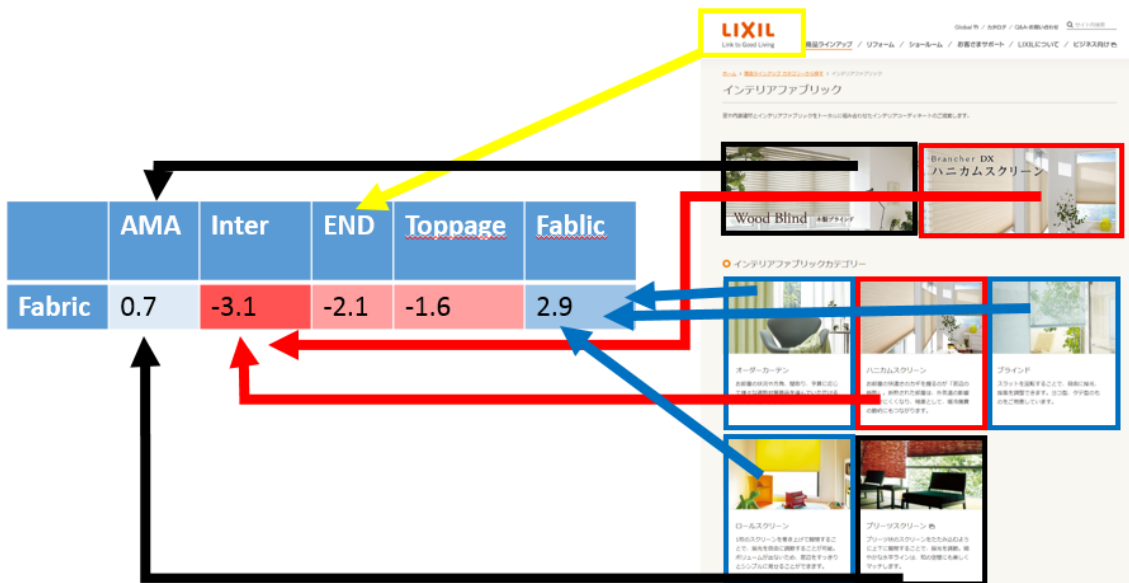


図 23 「Fabric」ページと推移確率の差分

6.2. コンバージョンブロックへの推移確率

休日平日間で「AMA」、「PRO」及び「Reform_C」の極限推移確率を比較した(図 24)。カタログはどちらも平日のコンバージョン率が高く、「PRO」では 4%も差があることが分かる。逆に「Reform_C」は 0.1%しか違いがない。

6.1 の推移結果から考えると、休日は施工主のアクセスが多くなると考えられるため、「AMA」に対するコンバージョン率が高くなるはずであるが、実際はそうになっていない。また、図 4 と図 5 を見ると、休日平日関係なく「Catalog」からの推移確率は「PRO」の方が大きい。しかし、その差分である図 18 では「Catalog」から「AMA」の推移確率は休日が高く、「PRO」は平日が高い。そのため、休日の閲覧者はカタログを余り見ず、商品ページを多く見ている可能性もある。

AMA	END
1.1%	-1.1%
PRO	END
4%	-4%
Reform_C	END
0.1%	-0.1%

図 24 各コンバージョンの差分

6.3. 滞留曲線

休日と平日の滞留曲線を重ねてみると、常に平日が下回る(図 25)。また、表 2 では閲覧者の離

脱率ごとに推移回数を比べてみた。これと 6.2 コンバージョンブロックへの推移確率からカタログコンバージョンについては平日のほうが効率的に行われている事が分かる。

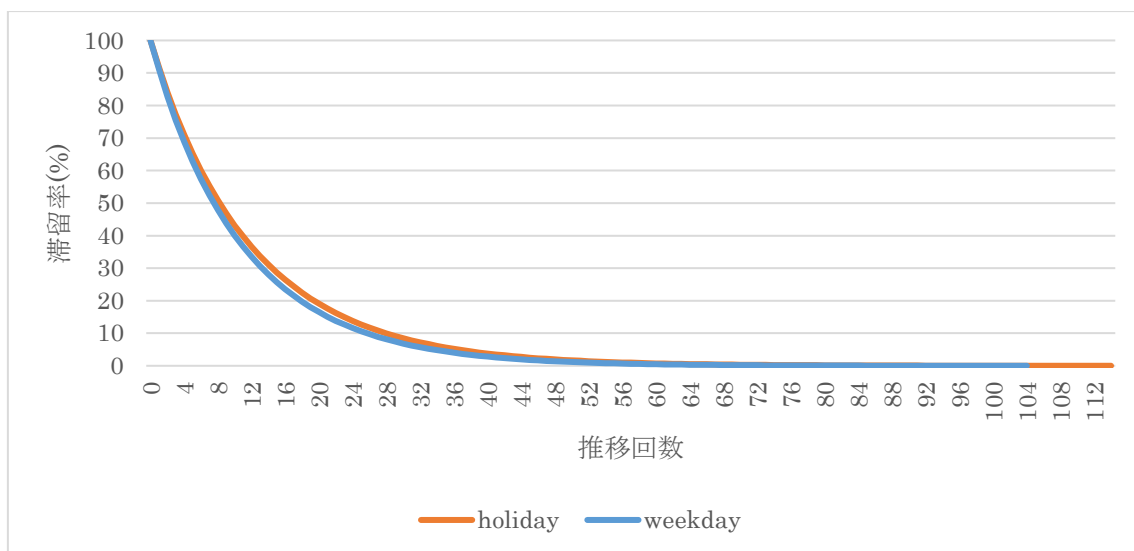


図 25 滞留曲線の比較

表 2 離脱率による比較

離脱率	休日(推移回数)	平日(推移回数)
25%	3	3
50%	8	7
75%	16	15
90%	27	25
99%	84	77
99.99%	113	103

6.4. 滞留中のコンバージョンブロックへの推移確率

休日平日間で滞留中於ける「AMA」、「PRO」への推移確率曲線を重ねる(図 26、図 27)。また、推移回数毎の推移確率を表 3、表 4 にまとめる。各グラフに於ける推移回数 6 回目程度までの急激な落ち込みは、表 2 で示した通り約 7、8 回で離脱率が 50%を超えていき、その後なだらかに離脱していくことに起因している。

推移確率の高い順では、平日の PRO:1.3%、平日の AMA:1.0%、休日の AMA:0.77%、休日の PRO:0.74%であり、平日では求められるカタログに差が存在しているが、休日では求められるカタログに差は大きく無い。これは休日平日間で閲覧者の性質が違う事を示していると考えられる。

また、単純に初回値の推移確率では、平日の「PRO」が他の場合に比べ非常に高い事が分かる。しかし、推移回数を重ねると、推移確率の値は、約 63%に収束するがこれは、他の場合よりも値が

小さい。これは初期状態分布において、「PRO」に推移する確率の高いブロックに多く閲覧者がいるからだと考えられる。

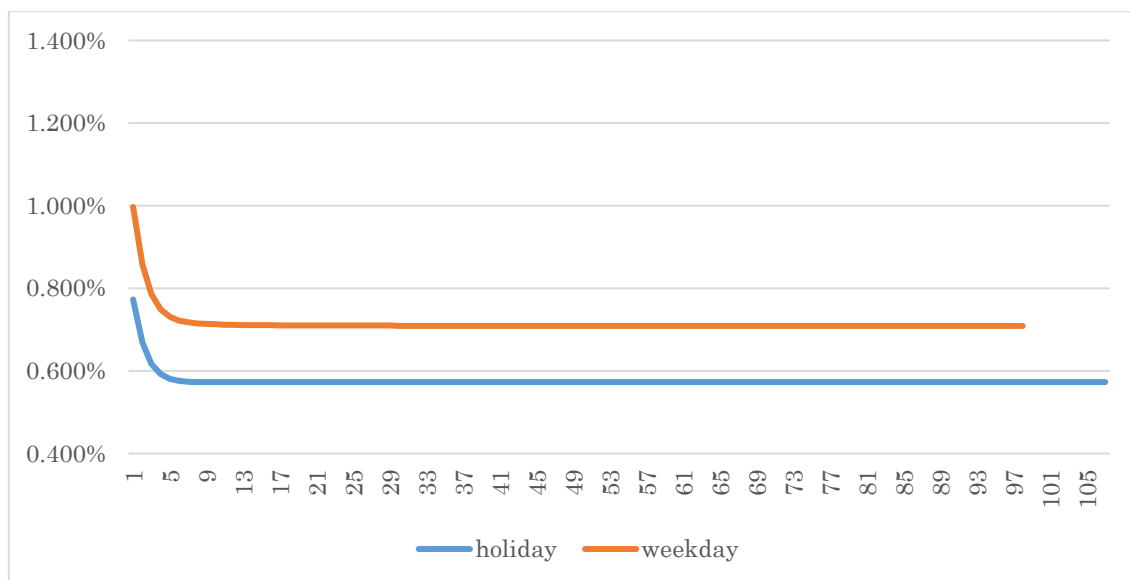


図 26 滞留中に於ける「AMA」への推移確率曲線の比較

表 3 「AMA」への推移確率曲線の推移回数による比較

推移回数	休日(推移確率)	平日(推移確率)
1	0.773%	0.997%
2	0.669%	0.858%
3	0.617%	0.786%
4	0.592%	0.749%
5	0.581%	0.731%
6	0.576%	0.722%
7	0.574%	0.718%
10	0.573%	0.714%
15	0.573%	0.711%

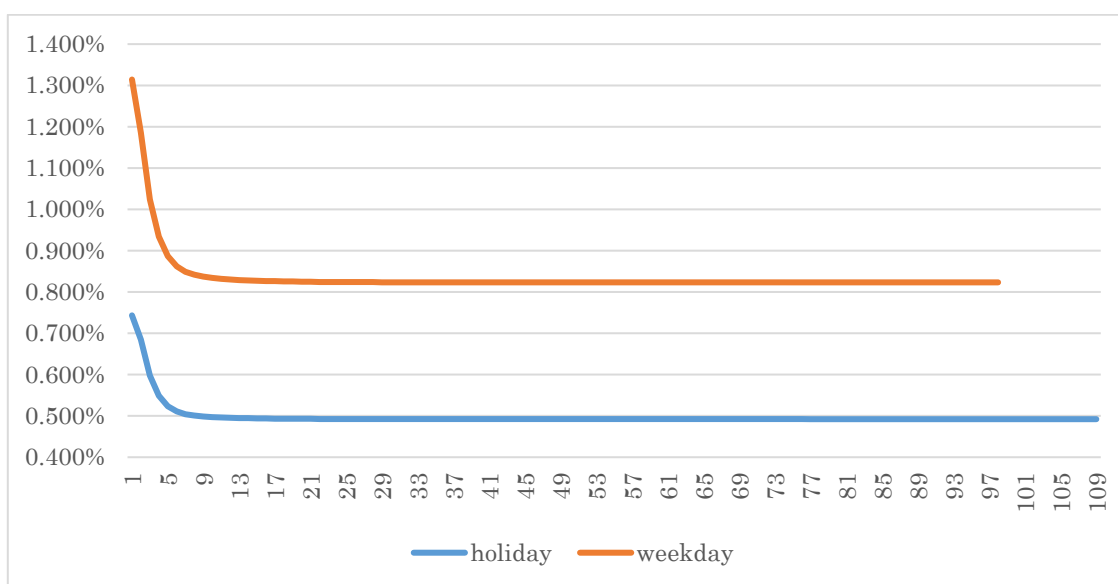


図 27 滞留中に於ける「PRO」への推移確率曲線の比較

表 4 「PRO」への推移確率曲線の推移回数による比較

推移回数	休日(推移確率)	平日(推移確率)
1	0.743%	1.314%
2	0.685%	1.185%
3	0.598%	1.024%
4	0.549%	0.934%
5	0.524%	0.887%
6	0.511%	0.862%
7	0.504%	0.849%
10	0.497%	0.837%
15	0.494%	0.827%

6.5. コンバージョン改善に向けた提案

上記の観察により、休日と平日では閲覧者のサイト内推移の違いがあることが判明した。しかし、2種類のカタログのコンバージョンブロックにおいて、休日平日間でブロック別の推移確率の差が逆転することなかった。例えば、「Window」、「Kitchen」、「Exter」では常に「AMA」がの推移確率が高く、「Public」、「Biz」ブロックでは常に「PRO」が高かった。これは休日平日に関係なく、そのブロックの閲覧者の属性が同じであることを示しているのではないかと考えられる。そこで今後は「AMA」への推移確率が高いブロックでは、施工主向けの情報を多く配置したり、「AMA」へのリンクを目に付く用に変更させる(「PRO」の推移確率が高いブロックでも同様に工夫する)ことで、コンバージョンへ誘導を自然に行う事が出来るのではないかと考える。

また、カタログについて、単純に施工主向けの情報として「AMA」、工務店向けの情報として「PRO」が存在しているが、休日平日どちらにおいても「AMA」から「PRO」への推移確率が高いことから、施工主はより詳しい、工務店向けの情報を知りたいと考えている可能性がある。そのため、施工主向けと工務店向けでカタログを分ける必要があるか今一度確認するべきだと考えられる。

7. 本手法の特徴と限界

7.1. アクセス解析としての評価

7.1.1. 他のアクセス解析手法との比較

他のアクセス解析手法がランディングページやコンバージョン、直帰率などの閲覧者の出入りに注目していることに対して、本手法では閲覧者のサイト内での回遊について注目をしている。回遊や回遊率に対しての概念的な Web サイトの作り方やアドバイスは非常に多く存在するが、アクセスログから回遊率そのものや滞留曲線そのものを可視化する手法一般的ではない。また、本手法を発展させ、コンバージョンブロックが何推移目でアクセスされたかを表示させる事が出来れば、他の解析手法とは大きく違ってくると思われる。

アクセスログを用いた閲覧者満足度や興味の解析方法ではサイトサーチアナリティクス [8]が主だったものであるため、本手法は実際の推移確率と想像していた推移確率の違いや、想像していなかった実際の推移確率を算出することで閲覧者の満足度や理解度、興味などを知ることが出来ると考える。

7.1.2. 他のアプリケーションとの比較

有名なアクセス解析ソフトに Google Analytics [5]がある。本手法において、Google Analytics に勝っている点は 1 点、異なっている点も 1 点だけある。

勝っている点は Google Analytics ではサイト内のブロック化(Google Analytics ではグループ化)は URL を基準に行っており、ブロック化したいページ群がディレクトリ構造になっていなければ一つにまとめることは出来ない。本手法ではそのような制限はないため、ディレクトリ構造でなくとも、ブロック化することが出来る。

異なっている点は、4.3.5 滞留曲線及び 4.3.6 滞留中のコンバージョンブロックへの推移確率の解析である。Google Analytics 単体では滞留曲線を表示させることはできず、そこから派生する推移確率も表示することが出来ない。

その他の点では Google Analytics には劣っている。本論文に於ける 4.3.4 コンバージョンブロックへの推移確率については、Google Analytics にはコンバージョンページへの閲覧者の流れを可視化することが出来る「ゴールフロー」機能(図 28)が備わっている。図 28 から分かるように、Google Analytics は非常に見やすく、数字の表示などの自由に設定することが出来る。

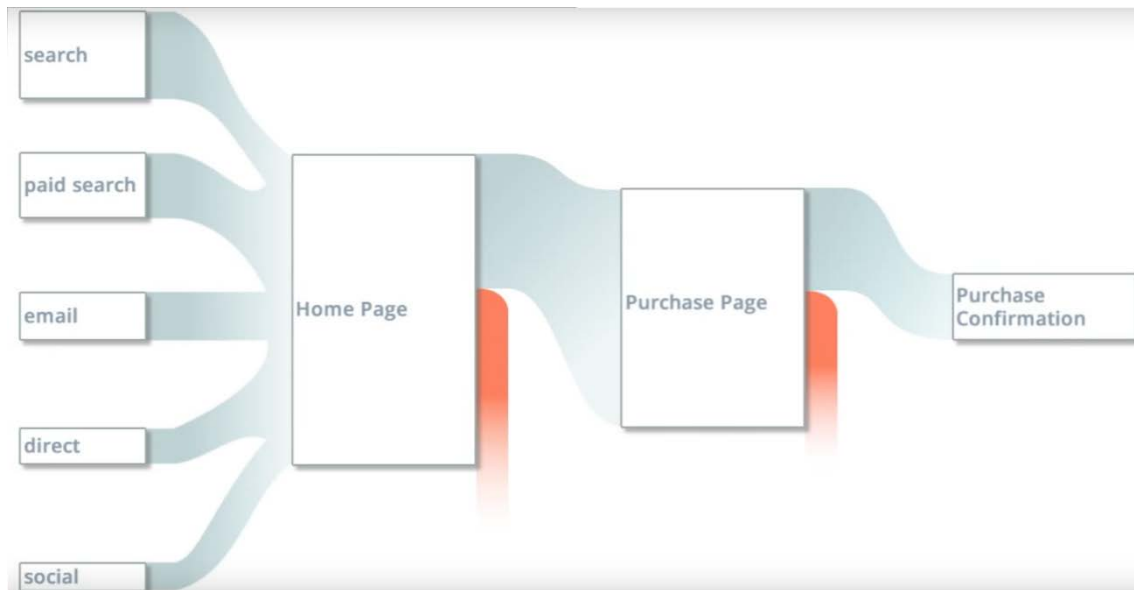


図 28 ゴールフロー [17]の動画の一部

7.2. マーケティングに於ける評価

アクセスログを使ったマーケティングでは、広告費などの兼ね合いで、「どうやってサイトの閲覧者を増やすか」「どうやって閲覧者を顧客にコンバージョンさせるか」が重視されている。結果、サイト内の構造がユーザーフレンドリーになっているか、閲覧者の欲しい情報がちゃんと存在しているかなどの閲覧者満足度についてはマーケティングではあまり触れられておらず、そのようなユーザーインターフェースについては Web デザイナーの仕事になっている。

このようにアクセスログが適切に使われず、そこから得られる情報が有効活用されていないため、マーケティング担当者や Web デザイナーが協力してサイト構造の改善やマーケティングを行うことが難しい。しかし、本手法を使うことで、ランディングページやコンバージョンブロックに限らず、サイト全体を俯瞰して行うことが出来るため、得られた閲覧者の推移情報をもとにサイト内の根本的な状態を確認し、上記したような改善を行うことが出来る。この解析手法により、マーケティング担当者はアクセスログに新しい価値を見出すことが出来る。

7.3. 手法の限界と課題

本手法から得られる知見は、情報源がアクセスログからのみであるため、多くが推測の上で成り立っている。得られた知見の正誤を確かめるためにも、他の情報(TVCM の時間帯、広告の種類、プロモーションしている製品など)を知る必要がある。そのため、今度はアクセスログだけではない情報を含めた解析手法提案する必要がある。

謝辞

本論文を書き上げるにあたり、主査である林高樹教授には本論文だけではなく、ゼミや授業での講義をはじめとした場面で、様々な事を教えていただきました。そのおかげで、統計学を知らなくはないと自信を持って言うことが出来るようになりました。

副査である大林厚臣教授、山本晶准教授には授業で様々な知見を教えていただいたからこそ、本論文を書き終えることが出来たと思っています。

X 社プロジェクトでは担当なさっている住田潮特任教授にも大変お世話になりました。また、今回のアクセスログのデータを提供いただいた X 社のインフォメーションエクセレンス部門の方々にもお世話になりました。たびたび私が送るデータに関する質問に快く応えていただきありがとうございました。

最後に大学院生活で自分に関わった全ての人々に、この場を借りて御礼を申し上げます。

参考文献

- [1] A. Bestavros, "Speculative data dissemination and service to reduce server load, network traffic and service time in distributed information systems," Proceedings of the Twelfth IEEE International Conference on Data Engineering, 1996.
- [2] I. Z. A. E. N. David W Albrecht, "Predicting User's requests on the WWW," 1999.
- [3] S. L. K. S. J. C. L. Alan L Montgomery, "Predicting online purchase conversion using web path analysis," Marketing Science, 2004.
- [4] R Development Core Team, "R: The R Project for Statistical Computing," R Development Core Team, 31 10 2016. [オンライン]. Available: <https://www.r-project.org/>. [アクセス日: 24 12 2016].
- [5] Google, "Google Analytics," Google, [オンライン]. Available: https://www.google.com/intl/ja_jp/analytics/. [アクセス日: 12 12 2016].
- [6] iannet, "Apache Log Viewer," iannet, [オンライン]. Available: <http://www.apacheviewer.com/>. [アクセス日: 12 12 2016].
- [7] Salvatore Sanfilippo, "Visitors," Salvatore Sanfilippo, [オンライン]. Available: http://www.hping.org/visitors/index_jp.php. [アクセス日: 12 12 2016].
- [8] L. Rosenfeld, サイトサーチアナリティクス アクセス解析と UX によるウェブサイトの分析・改善手法, 丸善出版, 2012.
- [9] Nexal, "上場企業における『Google Analytics』の導入状況調査 2015 年 12 月 78%の企業が Google Analytics を導入済み," 25 12 2015. [オンライン]. Available: <http://nexal.jp/blogs/20151225170217.html>. [アクセス日: 12 12 2016].
- [10] キーマンズネット, "企業におけるログの管理状況(2015 年)・前編," 26 1 2016. [オンライン]. Available: <http://www.keyman.or.jp/at/30008350/>. [アクセス日: 12 12 2016].
- [11] M. J. G.-M. D. Tak Woon Yan, "From user access patterns to dynamic hypertext linking," Computer Networks and ISDN Systems, 1996.
- [12] J. S. P. P. S. Y. Ming-Syan Chen, "Efficient data mining for path traversal patterns," IEEE Transactions on knowledge and data engineering, 1998.
- [13] 阿. 大草孝介, "アクセス遷移の傾向解析に基づくページ分類とそこから得られる知見について," 計算機統計学, 2014.
- [14] 佐. 康. 杉. 剛. 有. 雄. 田中 弦, アトリビューション 広告効果の考え方を根底から覆す新手法, インプレスジャパン, 2012.
- [15] 石井研二, 新版 アクセス解析の教科書 費用対効果がみえる Web マーケティング入門, 翔泳社, 2009.
- [16] アナリティクスアソシエーション, 新しいアナリティクスの教科書, インプレスジャパン, 2015.

- [17] G. Analytics, "Goals," 11 4 2013. [オンライン]. Available:
<https://www.youtube.com/watch?v=fMeKXsl7xT8>. [アクセス日: 4 1 2016].
- [18] C. O. Rachel Schutt, データサイエンス講義, オライリー・ジャパン, 2014.